Pergamon

0042-6989(95)00050-X

Vision Res. Vol. 35, No. 21, pp. 2985–3001, 1995

0042-6989/95 $9.50 + 0.00

Shape from Shaded Random Surfaces

ERIK DE HAAN,*† RODERIK G. F. ERENS,* ANDRÉ J. NOEST*

Received 11 July 1994; in revised form 3 February 1995

The perception of surface relief from random shading patterns is measured by having observers adjust three-dimensional local probes, the projections of which are superimposed on the image. Three observers perform four settings of 91 probes on each of 14 images. These images are generated by calculating the Lambertian reflectance of a random superposition of elliptical Gaussian hills and valleys illuminated by a single distant light source as well as by ambient light. Neither the surface reflectance equation nor the light source direction is conveyed to our observers in any way. Mathematically, this "pure" shape-from-shading problem has highly non-unique solutions. Perception of a well-defined, stable shape therefore implies that the ambiguity is resolved, i.e. a gauge is fixed. We analyse the surface ambiguity or gauge freedom which is left unconstrained by pure shading information and we investigate possible ways of restricting it. Statistical analysis of the curl component of the field of probe settings reveals that the settings are significantly consistent with an underlying perceived *surface*. In spite of the large theoretical ambiguity in the stimuli, the settings are reproducible and show considerable inter-observer agreement. Even the correlation of the settings with the real surfaces is surprisingly large. If the settings are compared to the real surface normals, one finds a series of *biases*, the strongest of which is that the global surface slant is systematically underestimated, even in those cases where ending occluding contours or high-contrast luminance ridges, indicative of "almost" contours, are present in the image. Another bias then is that the corresponding rims on the surface are seen as roughly parallel to the picture plane.

Shape from shading Luminance structure Depth Ambiguous figures Contour Correlation Occlusion

INTRODUCTION

Shading is one of the most powerful cues to depth structure, maybe not from the viewpoint of veridicality, but nonetheless from that of esthetic appeal and the *suggestion* of relief. In the development of the art of painting, shading played an important role in the emancipation of pictorial cues to depth structure and in artists' endeavours to apply these veridically, from the earliest beginnings.‡ Shading seems to enhance primarily information from other cues in the environment, such as occluding contours, motion parallax, stereo and prior information on the object shape (see, e.g. Bülthoff & Mallot, 1988, 1990; Todd, 1985; Rock, Shallo & Schwartz, 1978). This is not surprising, because shading information alone can never yield the correct depth structure uniquely. Multiple solutions are possible even in highly constrained situations (Horn, 1990).

Shading is fully specified by local surface reflectance properties and the angles between a surface element and the directions of light sources and the viewer. Thus, from a physical point of view, one may treat shading independently from other cues to depth, and even independently from other monocular luminance cues, such as aerial perspective, shadows and occlusion. Of course, this does not imply that the visual system also processes shading cues independently from others. Nevertheless, it is useful to know how our visual system performs in the ambiguous situation when only shading information is available. One wonders, for example, whether there will be ambiguity in the percept, as is the case with ambiguous drawings (Attneave, 1971), when there is so much ambiguity in the input. If not, how are the ambiguities resolved? The next stage then would be to ask how much objectively disambiguating stimulus information can be added before these internal biases are overcome.

In this investigation we are interested in the case of pure shading as generated by smooth complex "landscapes". Other cues, such as knowledge of light source or viewing direction and substantial knowledge of the object shape, are excluded as far as possible. The surfaces are generated by means of a random superposition of smooth elliptical hills and dales. We employ a paradigm in which observers set a probe (normal vector plus tangent plane) so that it "fits" the perceived three-dimensional orientation of local surface patches (see

*Utrecht Biophysics Research Institute, Faculty of Physics and Astronomy, Princetonplein 5, 3584 CC Utrecht, The Netherlands [*Email* e.dehaan@fys.ruu.nl].

†To whom all correspondence should be addressed.

‡Examples can be found already in Greek art (Pfuhl, 1923) and in the prelude to Renaissance art (Bunim, 1940). For a treatment of the concept of light and dark variations as a cue to depth over a more recent period of art history, see Verbraeken (1979).

Mingolla & Todd, 1986; Koenderink, van Doorn & Kappers, 1992).

When investigating shape from shading information on its own, it is interesting to consider the contributions of local and global properties separately.

Typically *global* properties involved are the direction of the light source, the average—i.e. global—slant and the orientation of the picture on the screen. We manipulate these independently. In different sessions, the main light source shines from above, from below and from the right. Moreover, two different inclinations of the light source from the viewing direction are investigated. Observers are never given information concerning the light source direction. Most surfaces are globally frontoparallel, but in some stimuli the surface is slanted in an upward direction, with slopes up to a maximum of 60 deg. Observers are not informed about the slant of any of the underlying landscapes. In the condition with 60 deg global slope, occluding contours occur, so we can compare performance with and without occlusion. In one case, an image is shown again in another condition, where it is rotated by 120 deg around the viewing direction (i.e. landscape and light source both undergo the rotation). As a final check on the influence of global properties, we employ an *aperture* of varying size, through which the stimulus is viewed.

We analyse the influence of many *local* properties upon the observers' settings. The local surface properties that we consider are, in increasing order of contact with the real surface: (0th) height, (1st) slant and tilt, (2nd) curvedness, shape index and orientation of the largest principal curvature and (3rd) the derivatives of both principal curvatures in the directions of both of these principal curvatures. For a definition of these properties, see Koenderink (1990). We choose to keep terms up to third order, because third-order terms are the first that can be expected to yield unambiguous information in the local luminance distribution (Koenderink & van Doorn, 1993).

Local image properties can also be defined in a geometrical way (Koenderink & van Doorn, 1987). If observers resolve the shape-from-shading problem in a purely local manner, one may expect their settings to depend on the local structure of the luminance distribution. Hence, we compare observers' settings with local image properties thereby checking how much of the variability in the observers' settings can be accounted for by the local luminance distribution. We compare the settings with luminance properties up to second order, which is the lowest order to yield unambiguous information on the surface shape if the light source direction is unknown (Koenderink & van Doorn, 1993).

METHODS

Stimulus

Random "landscapes". The surfaces of the landscapes were calculated according to the following procedure (see Fig. 1).

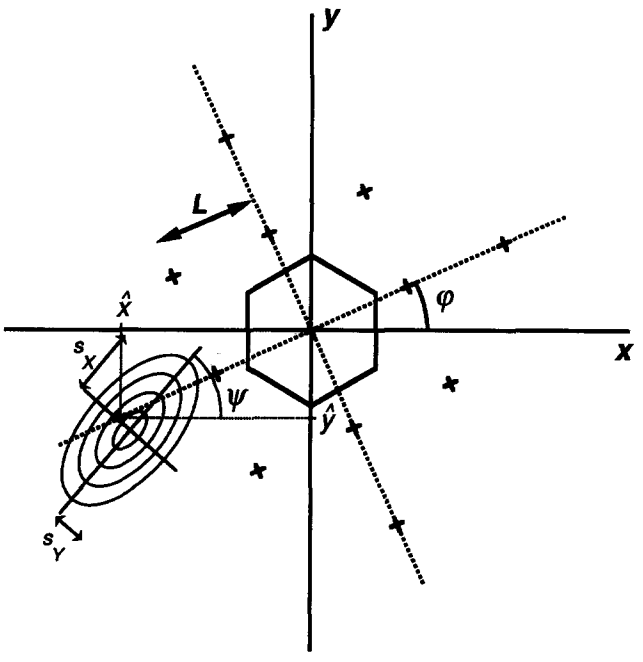

FIGURE 1. A schematic drawing indicating the construction of the landscapes employed in the experiments. Near each of the 13 points indicated by crosses a Gaussian hump is positioned. ϕ is the (random) orientation of the lattice of Gaussian hills and valleys with respect to the observer and L is the lattice distance. The size and shape of the triangulated area in which local measurements are performed is indicated by the hexagon. In the images where the landscape has been slanted with respect to the viewer's direction, this hexagon remains the same on the screen. One of the elliptical Gaussian humps, belonging to the lattice point $(\hat{x}, \hat{y})$, is indicated by some lines of constant height. It has a slight offset from the lattice point, an orientation ψ and spreads in orthogonal directions σ_X and σ_Y.

The 13 number pairs (k_1^i, k_2^i) with $|k_1^i| + |k_2^i| \leqslant 2$, indexed by i, define the positions $(\hat{x}, \hat{y})$ of a regular square lattice with lattice distance $L = 1.2$ deg (visual angle). This lattice has a random orientation φ in the vertical (x, y) plane:

$$\hat{x}^i = k_1^i L \cos\varphi + k_2^i L \sin\varphi$$

$$\hat{y}^i = k_2^i L \cos\varphi - k_1^i L \sin\varphi.$$

The surface $z(x, y)$ is constructed by superimposing elliptical Gaussian "humps" with random orientation ψ^i positioned near the 13 lattice points. Thus, performing the coordinate transformation

$$X^i = (x - \hat{x}^i - \delta_x^i)\cos\psi^i + (y - \hat{y}^i - \delta_y^i)\sin\psi^i$$

$$Y^i = (y - \hat{y}^i - \delta_y^i)\cos\psi^i - (x - \hat{x}^i - \delta_x^i)\sin\psi^i$$

for every Gaussian hump $G^i(x, y)$, the surface is defined as

$$z(x, y) \equiv \sum_{i=1}^{13} G^i(x, y) = \sum_{i=1}^{13} s^i \cdot A^i \times \exp\left\{-\frac{1}{2}\left[\left(\frac{X^i}{\sigma_X^i}\right)^2 + \left(\frac{Y^i}{\sigma_Y^i}\right)^2\right]\right\}.$$

The sign s^i, the amplitude A^i, both widths σ_X^i, σ_Y^i and the position offsets δ_x^i, δ_y^i of the humps were randomized uniformly over the following ranges: the sign

s^i was either + or −1 turning the hump into a hill or a valley respectively, the amplitude A^i. was within 0.86 deg ± 10%, the mean width $\bar{\sigma}^i \equiv (\sigma^i_X + \sigma^i_Y)/2$ was within 0.6 deg ± 10% and the location offset $|\delta^i|$ was <0.12 deg, which is 10% of the lattice distance. The *elongation* σ^i_X/σ^i_Y was within $0.6 \leqslant \sigma^i_X/\sigma^i_Y \leqslant 1.4$. The average overlap of the humps, determined by $\bar{\sigma}^i/L \simeq 0.5$, was rather large: enough to prevent the observers from noticing the build-up of the shading pattern from individual Gaussian humps. The z-value of the surface was calculated for 128 × 128 points $x, y \in [-7.8 \text{ deg}, 7.8 \text{ deg}]$. At every position the algorithm also determined higher-order properties of the local surface structure, up to third order. These properties were stored as far as possible in a Euclidean-invariant form, i.e. the local slant, tilt, shape index, curvedness, orientation of the second-order structure and the first-order derivatives of the principal curvatures in both principal directions (four third-order terms). For the definitions of these quantities, see Koenderink (1990).

Every observer saw a "standard" stimulus under four lighting conditions (see below). The stimuli were seen through a circular aperture of varying size. Outside the aperture the luminance was set to the average value of the shading pattern. The aperture has a diameter of 4.7 deg for most stimuli, but for the standard stimuli this diameter is also changed to 2.35 deg (slightly larger than 1.7 deg, the diameter of the triangulation, within which local measurements were performed) and 7.8 deg (showing the whole landscape: the aperture is now invisible to the viewer). In the case of the standard stimulus we also varied the global slant of the surface, introducing global slants of the z-axis of 20 deg, 40 deg and 60 deg respectively. For all slanted surfaces the global tilt was 90 deg; thus the surfaces were rotated about the horizontal (x-) axis. Lastly, we rotated the unslanted standard stimulus around the z-axis over an angle of 120 deg, which, with luminance azimuths of −120, 0 and 120 deg, led to the identical, albeit rotated, luminance distribution on the screen as one of the unrotated standard stimuli.

In addition to this set of 10 stimuli originating from the same landscape, each of the observers saw (a) two other stimuli generated by the above algorithm, (b) one stimulus without jitter in the elongation of the Gaussian humps, i.e. overlapping circular-symmetric hills and valleys and (c) a stimulus with Gaussian humps of half the amplitude (adding 10% jitter in amplitude, again). Table 1 shows exactly which image was viewed by which observer. Observers RE and SP were shown an almost identical set of stimuli, which makes more detailed comparisons between these observers possible. Observer EH viewed a partly different set of stimuli to check the influence of the specific choice of the parameters in the generation of the landscapes.

Generation of the shading pattern. The three-dimensional surface is illuminated by a unidirectional beam in combination with an ambient term from homogeneous background lighting. In combination with Lambertian reflectance of the surface, i.e. an isotropic

TABLE 1. Overview of the stimuli in the experiments

Image	Landscape	Light inclination	Light azimuth	Aperture size	Global slant	Observers	Additional remarks
1	1	40	0	0.6	0	RE, EH, SP	
2	1	40	120	0.6	0	RE, SP	
3	1	40	−120	0.6	0	RE, SP	
4	1	20	−120	0.6	0	RE, SP	
5	1	40	120	0.3	0	RE, SP	
6	1	40	0	1	0	RE, SP	
7	1	40	120	0.6	0	RE, SP	120 deg rotation of image 3
8	1	40	0	0.6	20	RE, SP	
9	1	40	−120	0.6	40	RE, SP	
10	1	40	0	0.6	60	RE, SP	
11	2	40	0	0.6	0	RE, EH, SP	
12	2	40	120	0.6	0	EH	
13	2	40	−120	0.6	0	EH	
14	2	20	0	0.6	0	EH	
15	2	40	−120	0.3	0	EH	
16	2	40	120	1	0	EH	
17	2	40	120	0.6	0	EH	120 deg rotation of image 13
18	2	40	−120	0.6	20	EH	
19	2	40	0	0.6	40	EH	
20	2	40	−120	0.6	60	EH	
21	3	40	120	0.6	0	RE	
22	4	40	120	0.6	0	EH	
23	5	40	120	0.6	0	SP	
24	6	40	120	0.6	0	RE, EH, SP	Circular-symmetric humps
25	7	40	120	0.6	0	RE, EH, SP	Half-amplitude humps

The inclination of the light source is the angle with the observer's direction or w-axis. The light source azimuth is an anti-clockwise rotation from the horizontal on the screen (u-axis). The slant is the angle of the z-axis of the surface with the w-axis, pointing from the screen in the viewer's direction. This angle is always directed vertically upward from the viewer, as the global tilt was kept at 90 deg.

reflectance that depends linearly on the size of a unit surface element as seen from the light source, these lighting conditions define the luminance at every point of the surface:

$$L(\mathbf{x}) \equiv \rho L_s \max[0, \mathbf{N}(\mathbf{x}) \cdot \mathbf{L}(\mathbf{x})] + \rho L_a,$$

where ρ is the surface albedo, L_s is the luminance of the light source, L_a the luminance of the ambient term, $\mathbf{N}(\mathbf{x})$ the surface normal at position $\mathbf{x}$, $\mathbf{L}(\mathbf{x})$ the light source direction at position $\mathbf{x}$. In our set-up we have $\rho L_s = 4.5\,\mathrm{cd/m^2}$ and $\rho L_a = 1.2\,\mathrm{cd/m^2}$. The light source has an inclination of 20 or 40 deg from the z-axis and an azimuth of -120, 0 or 120 deg from the horizontal (or x-axis). See Table 1 for precise specifications of the light source direction for every image.

The luminance distribution on the screen $L(u, v)$ is given by a simple orthogonal projection of the surface: $L(u, v) = L(x, y, z)$, with

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\sigma & \sin\sigma \\ 0 & -\sin\sigma & \cos\sigma \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix},$$

where σ is the global surface slant, which is a rotation around the horizontal x- or u-axis. The w-axis points in the viewer's direction. When the surface is slanted, the direction of the light source with respect to the surface remains identical; thus the shading pattern on the surface is not influenced by the global slant. Due to the projection, the pattern on the screen changes considerably. A few examples of the stimuli are shown in Fig. 2. In the case of a global slant of 60 deg [Fig. 2(c)], an algorithm performing hidden surface removal is employed at the projection stage.

In the course of an experimental session a depth probe visits 91 locations, which are on a regular hexagonal lattice with six points per side and a lattice distance of 0.17 deg. This lattice is fixed to the (u, v)-coordinates, i.e. the plane of the screen. At every location, the r.m.s. contrast and the local luminance derivatives up to second order are calculated, all being attenuated by a Gaussian window with a spread of 0.1 deg. This scale of 0.1 deg is sufficiently small, being below the dominant spatial frequencies in the images. The resulting values for the local luminance Taylor expansion form the so-called Gaussian derivative family which characterizes the local luminance up to second order (see, e.g. Koenderink & van Doorn, 1992). These are stored together with the observer's probe settings as far as possible in a Euclidean-invariant form. These local properties are calculated as follows. Defining the "fuzzy" derivatives of the image $L(u, v)$ at location (u_0, v_0):

$$\mathscr{D}_{mn}(u_0, v_0) \equiv \frac{1}{2\pi\sigma^2} \iint \mathrm{d}u\, \mathrm{d}v \left(\sigma \frac{\partial}{\partial u}\right)^m \left(\sigma \frac{\partial}{\partial v}\right)^n \exp\left[-\frac{1}{2\sigma^2}((u-u_0)^2 + (v-v_0)^2)\right] \cdot L(u, v),$$

with σ taken equal to 6.7 pixels (approx. 0.1 deg visual angle), we define the local luminance to be $\mathscr{D}_{00}$, the luminance *gradient* $\sqrt{\mathscr{D}_{10}^2 + \mathscr{D}_{01}^2}$ and the *orientation* of the gradient $\arctan(-\mathscr{D}_{01}/\mathscr{D}_{10})$. Next, with the $\hat{u}$-axis tangential to the isophote orientation and the $\hat{v}$-axis in the gradient direction, we represent the second-order structure invariantly, with (1) a term proportional to the isophote *curvature*:

$$\frac{\partial^2 L}{\partial \hat{u}\, \partial \hat{u}} \equiv \frac{\mathscr{D}_{01}^2 \mathscr{D}_{20} - 2\mathscr{D}_{01}\mathscr{D}_{10}\mathscr{D}_{11} + \mathscr{D}_{10}^2 \mathscr{D}_{02}}{\mathscr{D}_{10}^2 + \mathscr{D}_{01}^2};$$

(2) the change in gradient orthogonal to the isophote, which we call isophote *stretch*:

$$\frac{\partial^2 L}{\partial \hat{v}\, \partial \hat{v}} \equiv \frac{\mathscr{D}_{10}^2 \mathscr{D}_{20} + 2\mathscr{D}_{01}\mathscr{D}_{10}\mathscr{D}_{11} - \mathscr{D}_{01}^2 \mathscr{D}_{02}}{\mathscr{D}_{10}^2 + \mathscr{D}_{01}^2};$$

and (3) the change in gradient along the isophote, which we call isophote *splay*:

$$\frac{\partial^2 L}{\partial \hat{u}\, \partial \hat{v}} \equiv \frac{\mathscr{D}_{01}\mathscr{D}_{10}\mathscr{D}_{20} + (\mathscr{D}_{01}^2 - \mathscr{D}_{10}^2)\mathscr{D}_{11} - \mathscr{D}_{01}\mathscr{D}_{10}\mathscr{D}_{02}}{\mathscr{D}_{10}^2 + \mathscr{D}_{01}^2}.$$

Detection thresholds for these second-order luminance properties have been measured by Erens and de Haan (1993).

Experimental set-up

The stimuli were generated on an Apollo DN590 computer with the help of a graphical rendering package and displayed in 8-bit monochrome on a high-resolution colour monitor with dimensions 34.6×27.4 cm and 1280×1024 pixels. The colour look-up table was calibrated such that the screen luminance increased linearly as a function of the grey values. The experiments were done in a totally dark room. Observers viewed the screen monocularly from a distance of about 100 cm (small head movements were possible) with their dominant eye, which had normal or corrected-to-normal vision. They were asked to fixate loosely at the location on which they are required to report. Even though two of the authors, EH and RE, were observers and therefore knew about the design and purpose of the experiments, the fact that realizations were completely random made their *a priori* knowledge effectively useless. The third observer, SP, was unaware of the design and purpose of the experiments.

Psychophysical procedure

Every observer looked at 14 shading patterns (listed in Table 1), which were each viewed four times. The total of 56 sessions was conducted in random order, unknown to the observer. Observers did not have prior information on surface shape, light source direction, global slant etc. Observers were instructed to interpret all variations in luminance as arising from variations in relief, not in albedo.

After the image is put on the screen, the program displays the probe on the first of 91 randomly ordered locations, which lie on a regular hexagonal lattice with lattice distance 0.17 deg (visual angle). The probe consists of the orthogonal projection of a circle in three dimensions with an axle sticking out of the centre to one side, similar to the probe used by Koenderink *et al.* (1992). The probe is superimposed in dark blue upon the

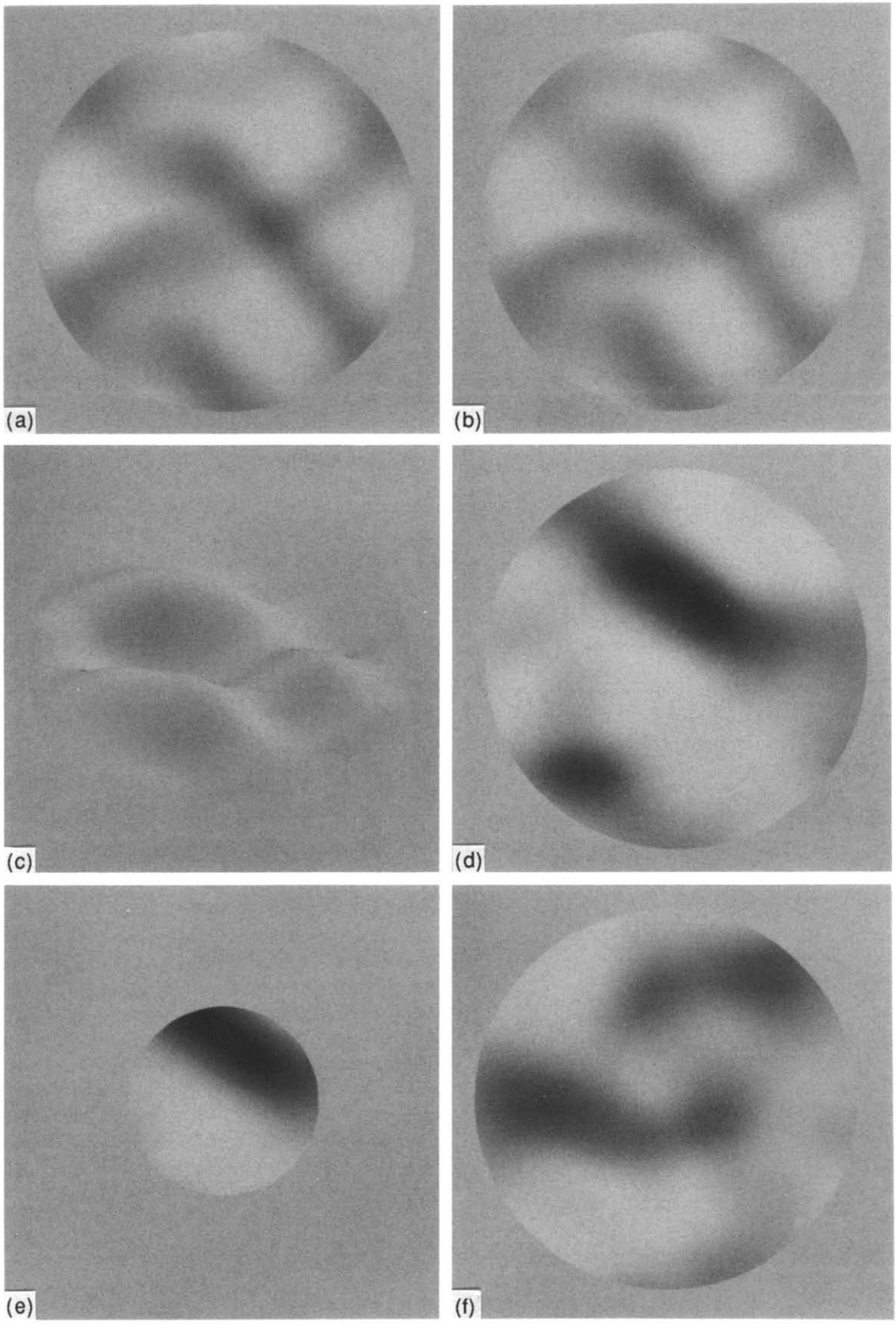

FIGURE 2. Some of the images that were viewed by the various observers. To obtain a strong depth percept the images are best viewed monocularly in a dark room, such as in the experiments. (a), (b) and (c) are the same landscape with increasing slant from the viewer. The slant of (a) is 0 deg, that of (b) 20 deg and that of (c) 60 deg. (a), (b) and (c) correspond to images 1, 8 and 10 in Table 1. These images have been generated with the same light source direction (azimuth, 0 deg; inclination from viewer, 40 deg). (d) The "standard" landscape of observer EH (image 13). (e) The same landscape, with the same light source direction and the smallest aperture (image 15). (f) Image 24, with circular-symmetric Gaussian humps.

grey shading pattern and has a diameter of 0.43 deg. It can be interpreted as a small "drawing-pin" attached to the three-dimensional surface. Observers are requested to adjust the probe with the help of the computer mouse, such that the circle looks as if it is tangential to the local surface, with the axle pointing in the outward normal direction. Observers can choose to (1) remove the probe temporarily from the screen so as to better view the local shading pattern, (2) unconfirm the previous measurement, or (3) confirm a setting that is satisfactory, thereby moving on to the next probe setting. One session of 91 lattice points took less than 15 min.

All observers and numerous visitors found the task of setting the probe both easy and natural. No observers required practice trials.

RESULTS

In total, four times 91 settings per image were measured, i.e. $4 \times 91 \times 14 = 5096$ three-dimensional unit vectors per observer. We analysed this dataset with the following questions and considerations in mind.

Even if the light source direction were known [but observers often misjudge the light source direction from shading information (cf. Pentland, 1982)] and the shading were taken to be Lambertian, there is still a two-parameter family of solutions of the shape-from-shading problem, parametrized by (1) the constant term in the luminance ρL_a, which cannot be extracted from the image because in our case there are no shadow zones on the object, and (2) the global gauge* on the tilts of the surface with respect to the light source direction, which is left free by the surface reflection equation.

It has been proposed that the natural way to analyse local surface orientation is in terms of individual slant and tilt parameters (Stevens, 1983). However, since slant and tilt with respect to the observer are not directly specified in the shading pattern (only the slant with respect to the light source direction is specified), we expect that the relation between slant and tilt settings and their veridical values would be ill-constrained. Indeed, as shown in Fig. 3, for a typical session for each of our observers, these relations are marred by very large scatter. Therefore we choose to analyse our results as far as possible in terms of surface normals. We deal separately with questions of consistency, reproducibility, inter-observer agreement and veridicality. We have to take into account the slant and tilt of the settings with respect to the observer only when investigating the influence of local conditions of surface and luminance structure on the settings.

At the coarsest level of analysis, we checked the distributions of surface slant and tilt with respect to viewing direction, and compared these with the slant and tilt distributions of the settings. The tilt distributions were roughly isotropic and did not differ significantly

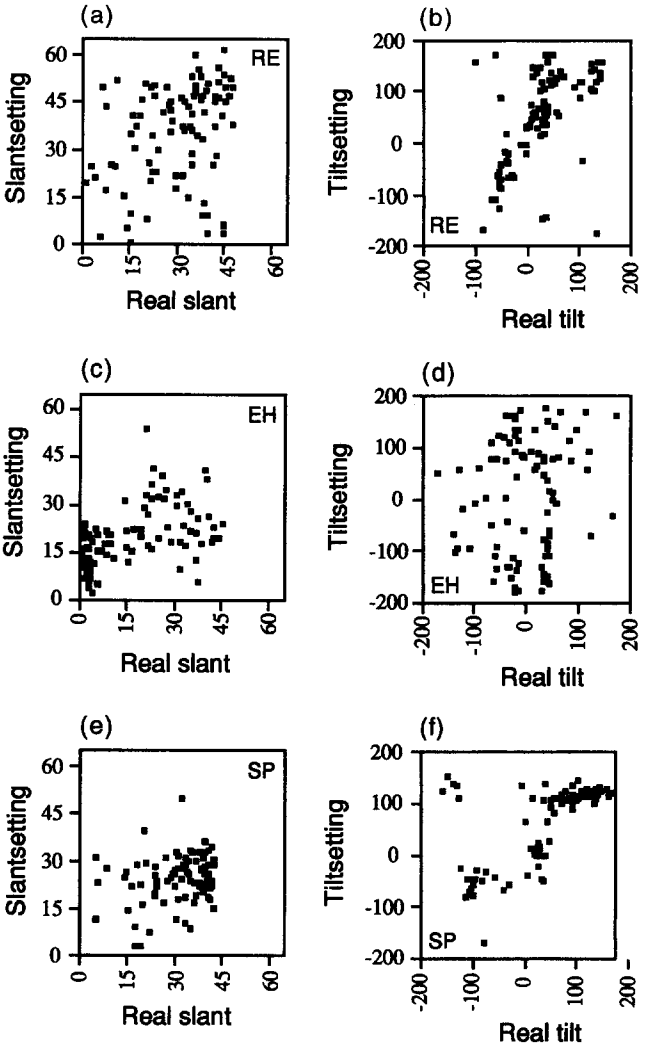

FIGURE 3. A comparison of all settings in a typical session to the real surface normals, for all observers. For all 91 points of a setting, the slant (tilt) as set by the observer is displayed as a function of the real slant (real tilt) at the same point. The setting of observer RE is his third setting on image 21, the setting of observer EH is his third setting on image 22 and the setting of observer SP is her third setting on image 23.

between settings and surfaces, but the slant distributions did. The spread in the slants was statistically significant, indicating depth ranges in the settings that deviate from the veridical depth ranges: 95% confidence intervals for the spread in the slants were 13.6–14.2 deg (real surfaces, RE); 16.7–17.4 deg (settings, RE); 14.4–15.0 deg (real surfaces, EH); 12.1–12.7 deg (settings, EH); 13.5–14.1 deg (real surfaces, SP); 11.2–11.7 deg (settings, SP). Thus, observers EH and SP show a compression of the depth range, whereas observer RE has an expanded range. We also noticed that the slant distributions for the settings were more skewed in the direction of the origin.

Consistency

A purely internal consistency check should be performed on the collection of normal-vector settings to see whether the settings are compatible with any surface at all. For smooth normal-vector fields, this amounts to checking whether the associated gradient vector field has

*At first sight, it would appear that the freedom in the tilts is local; in the Discussion the tilt freedom will be shown to be only global.

curl zero [is "conservative" or "integrable" (see Burke, 1985; Koenderink *et al.*, 1992)]. However, our sets of normal-vector settings do not form a dense vector field and therefore one may fit an infinity of surfaces that have these normals at the triangulation points. Thus, one must adopt a sufficiently restrictive surface *model* to be able to check the settings for consistency. Our model consists of a piecewise linear grad W field (W is the surface relief in the direction of the viewer). Thus, for each triangle we assume a quadratic surface ("Monge patch"),

$$W(u, v) \equiv au + bv + cu^2 + duv + ev^2,$$

suppressing the irrelevant constant term and imposing continuity in W and grad W across triangle edges. The three normal vectors around a triangle express 6 degrees of freedom; the surface $W(u, v)$ has 5 degrees of freedom, so in this case consistency fixes 1 degree of freedom in the settings. Indeed, for a triangle of vanishing size, this could be the vanishing of the *curl* of the vector field grad W: $W_{uv}(u, v) - W_{vu}(u, v) = d - d = 0$. For a quadratic surface, the line connecting two neighbouring vectors (i.e. the triangle edge projected onto the surface) is always a *parabola*. Thus, the tangent planes at two neighbouring points always meet at the (u, v)-value that is exactly at the center of the (u, v)-coordinates of the two points.

For every triangulation edge, we calculated the implied depth difference Δ_i over the edge i by projecting the two neighbouring settings on the plane spanned by the connecting edge vector and the w-direction. Then the depth difference is given by $\Delta_i = \frac{1}{2}E(\tan\phi_1 + \tan\phi_2)$ where ϕ_1 is the angle of the first projected vector with the w-axis, ϕ_2 is the angle of the second projected vector and E is the edge length. The *mismatch* M with our underlying model $W(u, v)$ is then the sum of the depth differences $\Sigma_{i=1}^{3}\Delta_i$ around the triangle. If the probes were set consistently, then $M = 0$ by Stokes' theorem (Burke, 1985), which equates the integral of the curl of grad W over any regular patch to the integrated height difference along its boundary.

For every observer and every edge in the settings, the spread in the depth difference settings, σ_Δ, is computed from the four settings for that edge. σ_Δ turns out to vary approximately linearly with the mean of the absolute value of the depth difference $|\bar{\Delta}|$ over the edge: regression $\sigma_\Delta = r|\bar{\Delta}|$ yields $r = 0.42$ ($R^2 = 77\%$, observer RE), $r = 0.36$ ($R^2 = 76\%$, observer EH) and $r = 0.38$ ($R^2 = 77\%$, observer SP). This approximate Weber-law behaviour for the mean absolute depth difference, which has also been found by Koenderink *et al.* (1992), may stem from inaccuracies in handling the two-dimensional probe. Approximate Weber-law behaviour has been found for the discrimination of the aspect ratio of an ellipse (Regan & Hamstra, 1992). In the case of our probe this would predict approximate linear dependence of the spreads in the depth differences on the mean absolute depth difference.

The consistency constraint $M = 0$ can be checked on closed paths of length n_e triangulation edges by assigning to each path a vector with components Δ_i in n_e-dimensional space, and analysing the variance structure of the distribution of these vectors. Both M and the Δ_i are found to be approximately normally distributed and a 99% confidence interval for the means of the Δ_i and M always contains 0. Ideally, all vectors should lie on the $(n_e - 1)$-dimensional hypersurface $M = 0$. Therefore, we check whether the variance per dimension of the vector distribution projected onto the $M = 0$ hypersurface is significantly larger than the (independent) variance of this distribution orthogonal to this surface. For any closed path the contribution per vector to the variance within the $M = 0$ hypersurface is:

$$v_\parallel = \sum_{i=1}^{n_e}\Delta_i^2 - \frac{M^2}{n_e}$$

[$(n_e - 1)$ degrees of freedom], and the variance orthogonal to the same hypersurface is:

$$v_\perp = \frac{M^2}{n_e}$$

(1 degree of freedom). Therefore, the appropriate F-ratio to test for a significant deviation from unity is, for N independent congruent closed paths l,

$$F[N(n_e - 1), N] = \frac{\dfrac{1}{N(n_e - 1)}\sum_{l=1}^{N} v_\parallel^l}{\dfrac{1}{N}\sum_{l=1}^{N} v_\perp^l}.$$

We perform the F-test both locally and globally. All images with global slant 40 or 60 deg, which yield different distributions for the depth differences and have depth discontinuities, are excluded from this analysis. So for every observer, results from 12 images or 48 sessions are taken into consideration.

(1) *Locally*, one has 27 independent triangles (i.e. not sharing any triangulation point) in the triangulation, so $N = 27 \times 48 = 1296$. Every triangle has $n_e = 3$ edges. The resulting F-ratios $F(2591, 1295)$ are 15.6 (RE), 16.1 (EH) and 13.4 (SP). These numbers are significantly different from 1, with vanishing P!

(2) *Globally*, one may take the border of the whole triangulation as a closed path (Stokes' theorem holds that the summed mismatch is equal to the mismatch over the border of the triangulation), such that $N = 48$. There are $n_e = 30$ border edges in the triangulation. The resulting F-ratios $F(1391, 47)$, with corresponding probabilities of finding at least this ratio in two samples from a normal distribution, are for RE: 1.8 ($P = 0.01$), for EH: 1.5 ($P = 0.07$) and for SP: 1.7 ($P = 0.03$). Although the effect is less prominent than in the case of local mismatches, there is significant global minimization of the mismatch at least for RE and SP.

Thus, we demonstrate that observers minimize the curl of the perceived depth gradients and set the probes so as to be biased towards consistency with *surfaces*. This result is complementary to that of Koenderink *et al.* (1992), which leaves open the possibility that the curl is not minimized within the collection of probe settings.

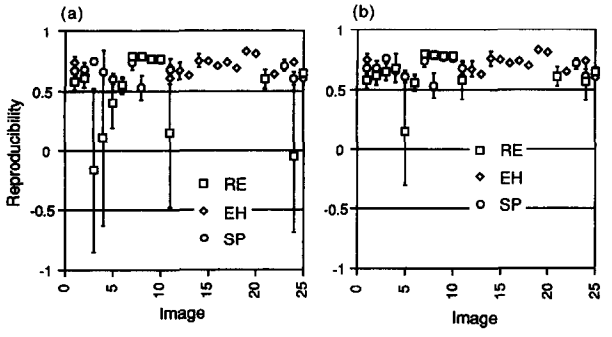

FIGURE 4. (a) Results of correlating the various settings on the same image, for all observers. The correlation coefficient for the 91 normals per session is defined in the text, as are the SDs in the coefficients. The smaller correlation coefficients are due to global inversions between settings, for observer RE. In (b) the inversions of RE are reinverted with respect to the true light source direction. Only the reproducibility of image 5 for observer RE remains small. This is the image with a small aperture size.

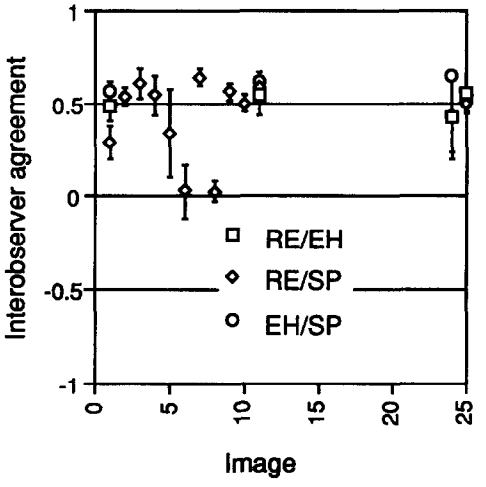

FIGURE 5. Correlation coefficients between observers for all images that are seen by at least two observers.

Reproducibility

The task we set our observers were unequivocal to them if the results were reproducible in different sessions. To check whether this was the case we calculated the correlation coefficient

$$C^{ij} \equiv \frac{\sum_\alpha (\mathcal{N}^i_\alpha - \bar{\mathcal{N}}^i)\cdot(\mathcal{N}^j_\alpha - \bar{\mathcal{N}}^j)}{\sqrt{\sum_\alpha (\mathcal{N}^i_\alpha - \bar{\mathcal{N}}^i)\cdot(\mathcal{N}^i_\alpha - \bar{\mathcal{N}}^i)}\sqrt{\sum_\alpha (\mathcal{N}^j_\alpha - \bar{\mathcal{N}}^j)\cdot(\mathcal{N}^j_\alpha - \bar{\mathcal{N}}^j)}}$$

between setting i and setting j. The summations are over the 91 triangulation points indexed by α, and the vectors $\mathcal{N}_\alpha$ are normals as set by the observer. With the four settings that were performed for every point, six different correlation coefficients can be calculated. We determined the standard deviation within this dataset. Note that not all correlation coefficients are independent, with the result that we get an underestimation of the spread. Results of this calculation of correlation coefficients are displayed in Fig. 4(a), for all three observers.

One notices that most settings do indeed correlate at coefficients around 0.7, except for five settings of observer RE, which also show large error bars. It turns out that these settings sometimes correlate and sometimes anti-correlate, due to global inversions which have long been known to be solutions of shape from shading (von Helmholtz, 1867). If the anti-correlating settings are inverted either in the w-axis or in the light source axis (which is the physically allowed solution), four of the five datapoints of observer RE below $C = 0.5$ are lifted to the strip around 0.7, as shown in Fig. 4(b). Here we have objective evidence that inversions occur with pure shading information. However, it is clear that the method is not sensitive enough to distinguish an inversion in the direction of the light source from an inversion in the viewer's direction. In Fig. 4(a), inversions between settings are relatively rare and occur only for observer RE: it seems likely that observers make presuppositions about probable directions of the light source (Brewster, 1826).

There is one datapoint in Fig. 4 for which an inversion does not help to make the settings correlate. This occurs at image 5, i.e. the condition with a small aperture. These settings are all weakly correlated [Fig. 4(a)]. Observer RE complained that it was hard to see the depth structure in this case.

In the calculation of other correlation coefficients (see below), we have first inverted the appropriate settings for observer RE with respect to the light source direction (inverted are two series of settings for images 3 and 24, and one series of settings for images 4, 5 and 11).

Inter-observer agreement

To determine the agreement of the settings between observers, we calculate correlation coefficients between a setting i of one observer and a setting j of another. In this case 16 different correlation coefficients can be calculated; standard deviations have been extracted from this dataset. The results are displayed in Fig. 5, which of course only incorporates those images that are shared

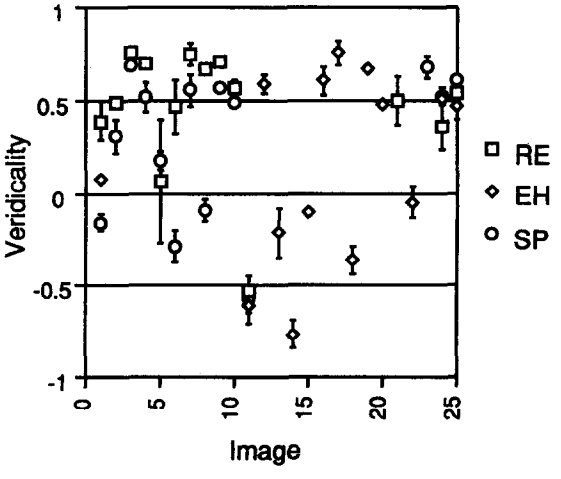

FIGURE 6. Correlation coefficients of all settings with the real surface normals at the triangulation points, for every observer.

by at least two observers. Even though most settings show a clear positive correlation, there are also settings which do not show any correlation, such as those for the images 6 and 8 for observers RE and SP.

Veridicality

The most straightforward way of checking the veridicality of the observers' settings is to correlate the normal settings with the real surface normals:

$$C^{iR} \equiv \frac{\sum_{\alpha}(\mathcal{N}^i_\alpha - \bar{\mathcal{N}}^i)\cdot(\mathbf{N}^R_\alpha - \bar{\mathbf{N}}^R)}{\sqrt{\sum_{\alpha}(\mathcal{N}^i_\alpha - \bar{\mathcal{N}}^i)\cdot(\mathcal{N}^i_\alpha - \bar{\mathcal{N}}^i)}\sqrt{\sum_{\alpha}(\mathbf{N}^R_\alpha - \bar{\mathbf{N}}^R)\cdot(\mathbf{N}^R_\alpha - \bar{\mathbf{N}}^R)}}.$$

Here, the deviation from the mean of a setting $\mathcal{N}^i$ is multiplied by the deviation from the mean of a real normal $\mathbf{N}^R$ at the same position. There are four correlation coefficients from which to determine the standard deviation. The results are displayed in Fig. 6, for all observers. Since higher correlation coefficients than those found in the test for reproducibility are not to be expected, the correlations in the figure, which are often >0.5, are encouraging. There are also some clear negative correlations (e.g. image 3 for RE and image 14 for EH). The settings on image 11, which was in the stimulus set of the three observers, are inverted with respect to the real surface for all observers (except for one setting of observer RE, which we have inverted before; see Fig. 4). The light source for image 11 has an azimuth 0 deg, not -120 deg where one expects more inversions because the light source is often assumed to be from above (see, e.g. Brewster, 1826). Lastly, there are a number of settings which do not have any significant correlation with the real surface. One notes that observer RE, who is most experienced in shape-from-shading experiments, always shows clear (anti-)correlations with the real surface (disregarding image 5, as discussed above).

A qualitative method to get an impression of the veridicality of the settings is to determine from the normals, as set by observers, a likely three-dimensional surface which matches the settings. This surface can then be compared visually with the original surface. We adopted the matrix inversion technique as described by Koenderink *et al.* (1992) to calculate a surface that best fits the settings according to a least-squares criterion. In effect, this reduces the 91×2 degrees of freedom of the probe settings to 91 depth values. Thus one expects this method of surface fitting to be somewhat robust against noise in the local normals. Figure 7 shows the results of an image with only a small correlation with the real surface (image 13 for observer EH) and an example of clear correlation with the real surface (image 24). The inversions in the data of observer RE for this image are also displayed.

Influence of local properties of surface and image

For every triangulation point, we store local Taylor expansions of the three-dimensional surface and of the luminance distribution. Thus we can easily check to which degree any of these local aspects of surface or image determines observers' settings.

Slant. We use the Spearman rank correlation coefficient to determine the dependence of the slant settings on the local parameters of surface and image. Although for the large dataset of 5096 settings the null hypothesis of independence is easily falsified (falsification with significance $P < 0.01$ occurs at correlations of $3.29/\sqrt{5095} = 0.046$), it is more interesting to look at the *amount* of correlation. We find correlations below 0.3 between any of the local parameters and all observers' slant settings. Generally, the strongest correlation occurs with the local surface slant (RE, 0.29; EH, 0.27; SP, 0.15) followed by that with the local luminance (RE, -0.10; EH, -0.28; SP, -0.14). In a mathematically correct shape-from-shading module large correlations with the parameters slant and luminance are expected. Correlations with other local parameters (including second-order luminance and third-order surface derivatives) are much smaller; only observer EH has correlations of about 0.2 with the local r.m.s. contrast and the second-order luminance properties "isophote curvature" and "isophote stretch".

This method does not rule out the possibility that some combination of the local parameters determines the response of the observers. Therefore, we apply multiple linear regression to the dependence of the slant settings on the parameters.* The local luminance derivatives up to second order can maximally account for 4.6% (RE), 16.9% (EH) and 6.7% (SP) of the slant settings' variability. The surface derivatives up to third order can account for 12.9% (RE), 11.4% (EH) and 6.6% (SP) of the slant settings' variability. All these percentages are so small that we can safely conclude that the observers do not perform the task purely locally.

Tilt. We calculate the dependence of the tilt settings on other parameters defined on a circle by means of the circular correlation coefficient (Fisher & Lee, 1983):

$$\mathscr{C}(\alpha, \beta) \equiv \frac{\sum_{i<j}^{N} \sin(\alpha_i - \alpha_j)\sin(\beta_i - \beta_j)}{\sqrt{\sum_{i<j}^{N} \sin^2(\alpha_i - \alpha_j)}\sqrt{\sum_{i<j}^{N} \sin^2(\beta_i - \beta_j)}},$$

between angular variables α and β for all N settings. If the tilt settings were to depend on the local first-order luminance structure, one would expect this dependence to be on the isophote direction, which is π-periodic, rather than on the gradient direction, which is (2π)-periodic. Therefore, we similarly calculate the measure $\mathscr{C}(\alpha, \beta)$ for the double angles. Correlations are very small: those between the tilts of the settings and the

*Since one should be careful with multiple regression when distributions are not strictly normal, we also apply multiple regression to the rank statistics by means of the Mann–Whitney–Wilcoxon procedure (Hettmansperger, 1984) with similar results. Consequently, we discuss only the results of ordinary multiple regression.

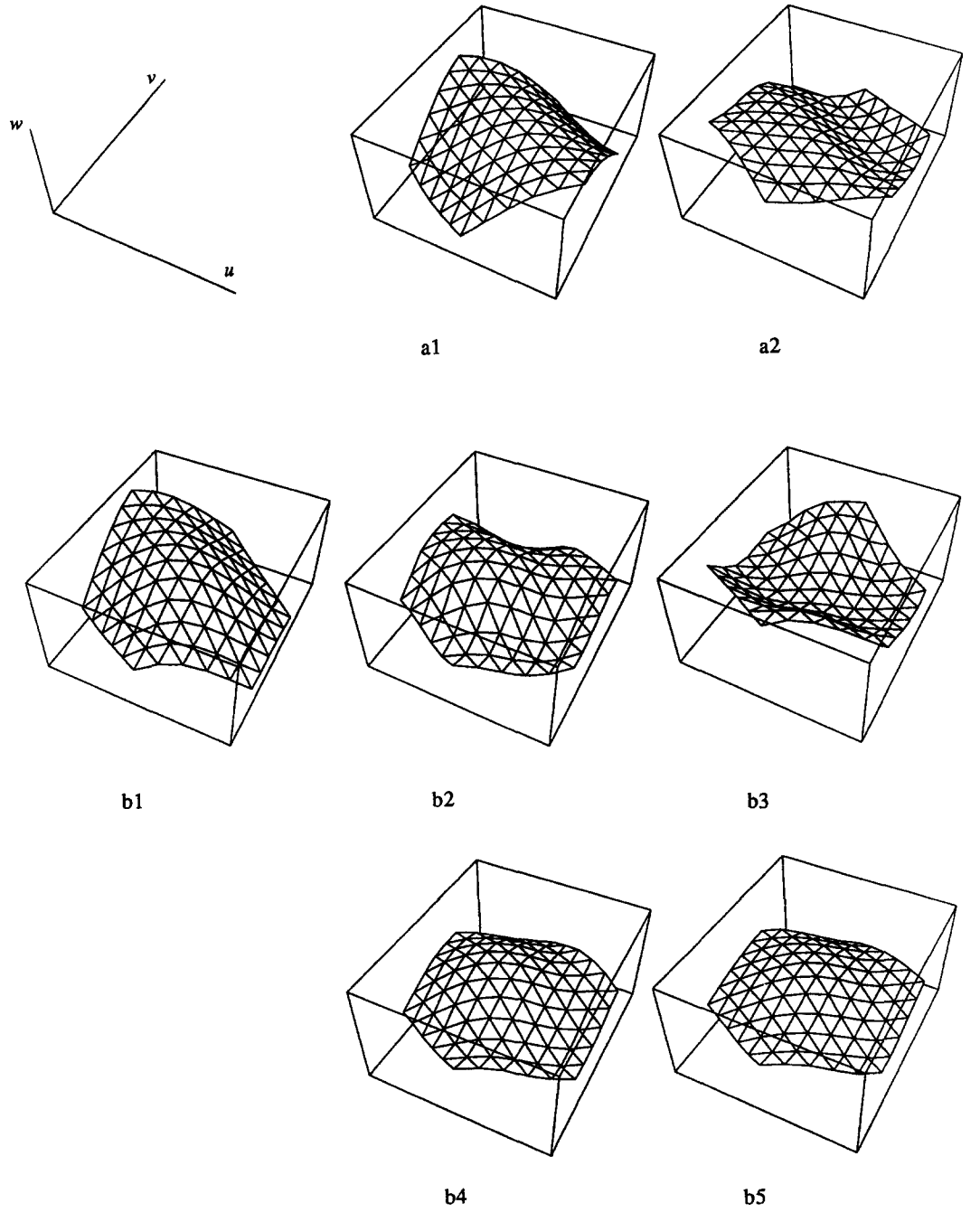

FIGURE 7. Examples of two of the surface shapes employed and some of the "perceived" surfaces as reconstructed from the field of normal vector settings. (a) An example of absence of correlation with the real surface. (a1) is the triangulated part of landscape 2, the surface from which image 13 is generated; (a2) fits best to the settings of observer EH for image 13. (b) An example of clear correlation with the real surface, also showing inter-observer agreement. (b1) is the measurement area of landscape 6, from which image 24 is generated, (b2) fits best to the first two settings of observer RE for image 24 and (b3) is the best fit to the last two settings of observer RE—note the inversion with respect to the surface at (b1). (b4) fits best to the settings of observer EH for image 24 and (b5) fits best to the settings of observer SP.

surface are around 0.02 and the double-angle correlations with the luminance gradient are between 0.02 (RE) and 0.08 (SP). Double-angle correlations between the local surface tilt and luminance gradient were below 0.03 for the collection of images of every observer. As expected, single-angle correlations are not larger than double-angle correlations. The double-angle correlation between the tilt settings and the orientation of the second-order surface structure is below 0.03 for every observer.

Second-order structure. For each triangulation point not on the boundary of the triangulated area, we fit a general quadratic surface $\mathcal{W}(u, v) \equiv au + bv + cu^2 + duv + ev^2$ to the settings at the central point and its six neighbours. Taking seven vectors instead of three (the minimum necessary to fit a second-order surface) allows a less noisy estimate of the perceived second-order structure. The fit is achieved by means of a stochastic-descent algorithm, which converges well and approximately reproduces the first-order structure of the setting

at the midpoint. The second-order properties of the fitted surface curvedness C, shape index S and orientation O of the largest principal curvature are calculated from $\mathcal{W}(u, v)$ and compared with the values for these parameters at the central surface point. Correlations between the second-order structure for the "perceived" and real surfaces are mostly small: Spearman rank correlation coefficients are for C, 0.18 (observer RE), 0.23 (observer EH) and 0.26 (observer SP); for S, -0.39 (observer RE), -0.2 (observer EH) and -0.2 (observer SP). Double-angle circular correlation C for O is 0.04 (observer RE), 0.01 (observer EH) and 0.03 (observer SP).

For all observers, the distribution of fitted shape indices contains more locally cylindrical surface shapes than are present on the surfaces. The shape index distributions are clearly bimodal for all observers, and peak near $|S| = 0.5$. If we take as a criterion for a shape to be cylindrical that $|S|$ is between $\frac{4}{9}$ and $\frac{5}{9}$, then the 90% confidence intervals for the fractions of locally cylindrical patches are for observer RE, $14 \pm 1\%$ (real surfaces), $17 \pm 1\%$ (settings); for observer EH, $17 \pm 1\%$ (real surfaces), $19 \pm 1\%$ (settings); for observer SP, $15 \pm 1\%$ (real surfaces), $20 \pm 1\%$ (settings).

Influence of global properties of surface and image

The global properties of the images are not of great influence upon the settings. Figures 4 and 6 illustrate that images which differ in the size of the aperture, position of the light source, global slant and global orientation, do not yield marked and consistent differences in the settings. For example, inversions among settings [Fig. 4(a)] occur for observer RE for all directions of the light source.

Also, in Fig. 6, no single global variable seems responsible for the repeated inversions with respect to the real surface. Interestingly, image 11, with a light source azimuth of 0 deg, shows a highly negative correlation with respect to the real surface for all observers. The only cases where an image with a light source azimuth of 120 deg—the direction frequently used by cartographers and painters—is clearly inverted, are two settings of observer RE for image 24.

From Fig. 4 one infers that only for observer RE is the condition with a small aperture substantially more difficult than the conditions for other apertures. We have calculated the correlation coefficients between the settings for every pair of images that differ only in the size of the aperture. These correlation coefficients remain above 0.5 for all observers, except for the correlation coefficient between the settings of observer RE on images 2 and 5. The veridicality of the settings always increases with increasing aperture size when compared with the settings on images with the same light source direction, though the increase is not always significant (Fig. 6).

We have also calculated correlation coefficients between settings for images that differ only in the direction of the point light source. If only the light source *inclination* was different all these correlation coefficients remain above 0.6. This indicates that the *contrast* in the image is not very influential on the shape-from-shading performance—a finding corroborated by the high correlations with the real surface of the image with half amplitude hills and dales and thus lower contrasts (image 25) (cf. Fig. 6). On the other hand, the light source *azimuth* has a large influence upon the settings: all correlation coefficients between settings on images that differ only in light source azimuth are <0.4; in fact, the correlation coefficient is even below 0.15 in six out of nine cases!

We adopt the following approach to reconstruct the assumed direction of the light source from the settings:* in a shading pattern the luminance varies proportionally to the cosine of the slant *with respect to the light source*. Therefore, if observers assume a given light source direction, one may expect variability in the slant settings with respect to this direction to be smaller than variability in the slant settings with respect to other arbitrary vectors. By means of a stochastic-descent algorithm we determine the direction of the light source that minimizes the slant variability with respect to that direction. This was done for every image, for which variances are determined from the set of four settings. The algorithm always converges at inclinations of the light source direction from the viewer's direction which are near 90 deg. This can be attributed to Weber's law for the spread of the settings as a function of slant with respect to the observer, together with the small tilt variability found (see also Koenderink *et al.*, 1992). Therefore, we take the inclination of the light source equal to the real light source inclination and run the program to determine only the light source azimuth that minimizes the variation of the slant with respect to the light source. The algorithm converges quickly. Results are shown in Table 2. Clearly, the algorithm converges often at azimuths around 0 deg ("3 o'clock"), and also frequently at azimuths around 135 deg, for every observer: of the 42 values for the best-matching azimuthal lighting directions in the table, 18 are in the interval [-30 deg, 30 deg] and another 11 are in the interval [130 deg, 160 deg]. All the other values are near these two intervals, except for one occurrence of the lower-left direction (-152 deg for observer SP at image 23). The precedence of two assumed lighting directions occurs for all observers, though observer SP has only two occurrences in the interval [130 deg, 160 deg] and somewhat more around 0 deg (lighting from the right).

For the pairs of images that are rotated copies (images 3 and 7, as well as 13 and 17), we calculated correlation coefficients between the settings on corresponding points on the surfaces. In this case, the observers' settings on images that are rotated back over $+120$ deg are rotated over -120 deg before the correlation. The resulting correlation coefficients are 0.66 for observer RE, -0.06 for observer EH and 0.59 for SP. The correlations for SP and RE are of the same order as has been found for

*But note that this approach is only valid *if* there is a relatively stable assumed direction of the light source.

TABLE 2. Results of a stochastic-descent algorithm determining the azimuthal direction of the light source that best matches the settings

		Azimuthal direction best matching the settings		
Image	Real azimuth	Observer RE	Observer EH	Observer SP
1	0	28	146	154
2	120	33	—	14
3	−120	134	—	−14
4	−120	133	—	15
5	120	−2	—	131
6	0	30	—	−17
7	120	64	—	72
8	0	153	—	−6
9	−120	114	—	1
10	0	42	—	6
11	0	120	143	−9
12	120	—	10	—
13	−120	—	−33	—
14	0	—	151	—
15	−120	—	4	—
16	120	—	7	—
17	120	—	153	—
18	−120	—	9	—
19	0	—	148	—
20	−120	—	−16	—
21	120	70	—	—
22	120	—	175	—
23	120	—	—	−152
24	120	10	166	49
25	120	72	−13	78

As only the local surface slant with respect to the light source is specified in a Lambertian shading pattern, one may assume the variability of the adjusted slants with respect to the assumed light source direction to be minimal. The algorithm calculates the mean slant variance of the settings at the 91 triangulation points with respect to a random (light) vector and then minimizes such a mean slant variance over the azimuthal direction of the vector. The inclination of the vector is always taken equal to the real light source inclination.

reproducibility (Fig. 4). The fact that *positive* correlations are found for all settings of SP and two settings of RE (image 3 resulted in inversions between settings for observer RE) is rather surprising, in view of the change of light source azimuth from downward to upward. If an upward direction were assumed by observers one would expect negative correlation coefficients. The absence of correlation for observer EH may be related to the absence of correlation of his settings for image 13 with the real surface normals (cf. Fig. 6). If one looks at the inferred lighting azimuth for observer EH's results for the two rotated images 13 and 17 (Table 2), one can see that EH probably makes an approx. 90 deg misjudgement of the light source azimuth for image 13, whereas the light source azimuth for image 17 is estimated much better.

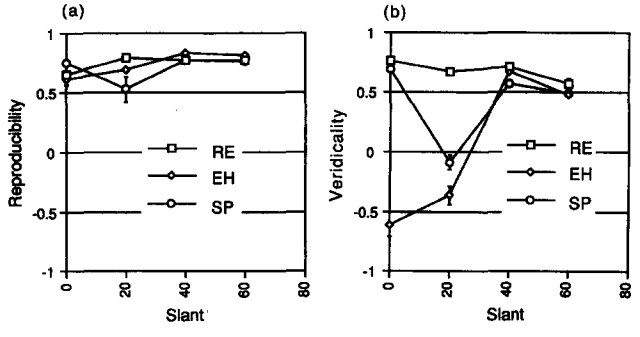

FIGURE 8. (a) Reproducibility of the settings on the same landscape with increasing global surface slant. For observers RE and SP, results are shown for landscape 1 and images 3, 8, 9 and 10. For observer EH, results are shown for landscape 2 and images 11, 18, 19 and 20. Note that for these images the light source azimuth may differ. The azimuthal lighting direction is either 0 or −120 deg (see Table 1). These results are also shown in Fig. 4. (b) Correlations of the same settings as in (a) with the real surface, as a function of the global slant of the real surface. These results are also shown in Fig. 6.

The results with slanted surfaces replicate results obtained by Reichel and Todd (1990), namely that the global slants are biased in the upper vertical direction (in our case the veridical slant direction), although in our results occluding contours do not serve to increase the amount of veridicality (cf. Todd & Reichel, 1989). We replot the results of Figs 4 and 6 as a function of the mean surface slant in Fig. 8. Clearly, there is no influence of the global slant of the surface upon reproducibility. As shown in Fig. 8(b), there is no general trend that increasing slant increases the veridicality of the settings (though observer SP has one "outlier" at slant 20 deg). Only the inversions of observer EH with respect to the real surface are undone when ("almost") contours are present in the image. These results are surprising: the slanted surfaces subjectively provide a more

"convincing" depth *impression*. If one looks at Fig. 2(c) the 60 deg slanted surface for observers SP and RE, one notices occluding contours which yield a vivid impression of depth. Although strictly *occluding* contours did not occur in the other experiments, "almost" contours are still created in the shading pattern of the less slanted images in the standard series. On a surface patch with a locally extreme slant towards the observer there will be a relatively steep ridge in the local luminance distribution.

When investigating the relief perceived in monocularly viewed pictures, one should carefully distinguish structure *in the frontoparallel image plane* (u, v-coordinates) from *relief along the observer's axis* (w-coordinate). Sharp luminance edges or ridges provide mainly clear structure in the image plane. This structure may be interpreted as occluding contours, which does not add much to the depth structure proper: which side of the contour is the foreground, which the background? What is the orientation, in three dimensions, of the corresponding *rim* on the surface? In fact, we find (see Fig. 8) that observers resolve this ambiguity in a clearly biased way for the roughly horizontal contours in our images 9, 10, 19 and 20. The nature of this bias may be appreciated by looking at Fig. 2(c). *Subjectively* the part of the image that is below the contour appears to lie in the foreground (see also Reichel & Todd, 1990). In the case of our stimuli this resolution of ambiguity happens to be veridical, but things could have been different: try looking at Fig. 2(c) upside down. The results displayed in Fig. 8 show that no inversions with respect to the real surface are registered for slants >20 deg.

If the depth ordering about a contour is known, the sign of curvature of the contour yields the sign of the local Gaussian curvature of the surface (Koenderink, 1984). To put it differently: normal vectors in the foreground near an occluding contour cannot be oriented away from the contour, in any projection. To check the settings near contours, we look at all settings that have a slant higher than 70 deg. These are 65 settings all near a contour: 28 from observer RE, 24 from EH and 13 from SP. If for this set of data points we determine the circular correlation coefficient $\mathscr{C}$ between (1) the tilt of the setting and the real tilt, (2) the tilt of the setting and the luminance gradient and (3) the real tilt and the luminance gradient, we find the following values: 0.71 (tilt setting and real tilt), 0.82 (tilt setting and gradient) and 0.59 (real tilt and gradient). These numbers are much higher than correlations reported in the previous section between these parameters for the whole dataset, which indicates that observers are indeed guided by the occluding contours to set the tilt of the probe. The correlation between tilt and luminance gradient is the highest of these values and is especially higher than the correlation between the real tilt and the gradient in the images. Thus, the vectors are set roughly orthogonal to the contour not only on the contour itself, where this is correct, but also on a finite region on the foreground, where this is generally incorrect. Such a bias implies that observers perceive the rim as if it were roughly parallel to the picture plane!* Admittedly, this observation is as yet based on few data (65 settings), but as a first quantitative measurement of the perceived direction of the rim in the case of pure shading information, we think it is valuable.

*The assumption that the original of a contour in a picture is a planar space curve has been considered before by Stevens (1981). Stevens argues that such an assumption is likely to be made when humans interpret line drawings.

DISCUSSION

Summary of results

Analysis of the probe settings provides objective evidence for the notion that random smooth shading patterns generally induce a stable three-dimensional shape percept. This is surprising because the shading pattern itself does not determine a unique shape. Moreover, the period over which the entire series of settings was performed consisted of more than a month for all observers. The only cases for which multiple shapes are found are five images of observer RE: the different shapes turn out to be related by simple global inversions (Fig. 4). Although the presence of inversions within the set of RE's settings sets observer RE somewhat apart from the other observers, one can generally conclude that all three observers' datasets show the same features. This is especially interesting in the case of observer EH, who viewed a stimulus set which was largely different from that of the other observers (see Table 1). From the fact that EH shows qualitatively the same results as RE and SP, we conclude that the randomness in the underlying surfaces is large enough to prevent the results from being dominated by fortuitous features of the surfaces. Almost all images induce reproducible settings ($C \simeq 0.7$). The (anti-)correlation of the settings between observers and with the underlying real surface generally lies in the range 0.5–0.7, but there are some exceptional cases of very low correlation (Figs 5 and 6).

Neither purely local properties of the shading pattern nor very global characteristics of the stimulus lead to a straightforward explanation of the settings. Of the global aspects that we varied, only the *azimuth* of the light source direction influences perceived shape so strongly that consistently small correlations are found between probe settings on the same surface, when only the azimuth of the light source differs by 120 deg. However, the *inclination* of the light source from the viewing axis does not matter very much.

The presence of occluding contours in some images does not lead to more veridical settings (Fig. 8), although it does strongly enhance the subjective impression of depth structure.

Ecological optics of shape from shading

It is important to stress again that in ecological optics there is a multi-parameter family of permissible interpretations of a smooth luminance pattern, such as the patterns we employed. The solutions range from a flat

painted surface (or, indeed, a luminescent display tube) to the rather far-fetched possibility of the screen luminance having been determined by aerial perspective in front of a dark background, in which case luminance would covary with distance from the observer. Here, we consider only interpretations that treat the luminance pattern on the screen as originating from shading only. The question is then what presuppositions are needed to arrive at a unique shape-from-shading percept?

Firstly, one should interpret all variations in the luminance as arising from variations in the surface normals. This was also the main instruction given to our observers. Whether observers see variations in surface albedo and incorporate them in their settings cannot be extracted from our data. If one assumes a fixed, monotonic relation between the local luminance in the image and the slant of the local surface normal with respect to the light source (a very general assumption in the case of stationary viewing), there is only a two-dimensional *global* gauge freedom left. One can see this by returning to the local surface model that we adopted to investigate the consistency of the local settings. The proposed quadratic fit to the settings becomes exact for small enough triangles: locally, the complex shading pattern becomes a gradient and the complex underlying landscape becomes quadratic. This implies 5 degrees of freedom in the surface within the first triangle (*v.s.* subsection Consistency). The shading pattern provides 3 degrees of freedom (3 luminances) at the vertices, so 2 degrees of freedom are left unconstrained. Adding a fourth point to create a new triangle adds only 1 surface degree of freedom, because every triangle has its own "curl = 0" constraint. The single degree of freedom is immediately fixed by the value of the luminance at the fourth point.* The same reasoning applies to all further additions of triangles that are adjacent to those already accounted for. Thus, the 2 degrees of freedom that are left unconstrained in the first triangle are *global*, i.e. fixing them at one point fixes them at all the other points.

Experimentally, we find reproducible surface normal settings, indicating that the gauges are indeed fixed by the observers. How can the visual system proceed to fix the two global gauges? We distinguish the following proposals, which we compare with our experimental results.

- Horn (1990) proposes to use the knowledge about the local surface orientation at special loci, such as on contours. However, it should be noted that propagating this constraint to the correct foreground shape is hindered by the ambiguity about which side of the contour belongs to the foreground, and by the reflectance properties being highly uncertain near occluding contours. In most of our images there are no special loci where the real surface normal is known. Occluding contours occur in some of our images and although we find some bias for the tilt to be set perpendicular to the contour, there are no substantial improvements in veridicality compared to the images without contours (Fig. 8). Even though occluding contours contribute strongly to the perception of depth relief, the veridicality of this contribution should not be overestimated: even if the depth order is determined and the sign of the Gaussian curvature is known locally, a misperception of the orientation of the surface rim leads to large errors in local slant and tilt estimates. Therefore, our results plead against the use of special loci to fix the global gauges.
- Another point of view that is often adopted (starting with Brewster, 1826) is that the visual system *assumes* a likely direction of the light source. Lighting from above is then mostly taken to be the best candidate. Note that in ordinary viewing this would often lead to errors, e.g. when an object is held in the hand the light often comes from *below* with respect to the surface. Rittenhouse (1786) observed that if the light source direction on the surface is inverted by means of a mirror, a perceived inversion of the surface is undone, indicating that observers can discount a perceived illuminant direction. In our set-up the light source direction is not given to observers, so it is important to check what light source direction is assumed.

Our results do not confirm the expectation of one *unchanging* assumed direction of the light source,

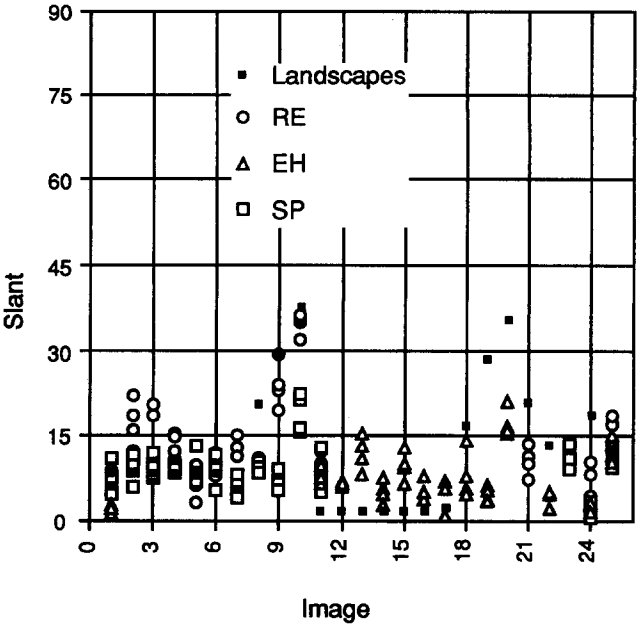

FIGURE 9. The slant of the average normal vector from the 91 settings in every experimental session. In addition, the slant of the average normal vector of the real surfaces is shown (■). Clearly, this "perceived global slant" was usually small. Only when the real surface has a large slant (images 9, 10, 19 and 20), does the perceived global slant increase somewhat, especially for observer RE.

*Of course, it may occur that the gauge in the first triangle is chosen such that in some other triangle it becomes impossible to fit a second-order surface, given the local luminance. This circumstance further restricts the space of solutions but does not reduce the dimensionality of the gauge freedom. This is precisely the reason why one expects better performance in the case of stimuli with increased aperture size, even if all apertures include the triangulation area.

because inversions with respect to the real surface are found for all lighting directions employed (Fig. 6). However, from the results of our stochastic-descent algorithm that infers a perceived lighting direction from the settings (see Table 2), we have concluded that the azimuths of the reconstructed lighting directions cluster within two regions, namely around 3 o'clock and around 10 o'clock on the monitor, for all observers. This is consistent with Brewster's (1826) proposal of an assumed light source direction.

- In ordinary viewing the underdetermination of the shading cue to depth structure may be resolved by exploiting the various other depth cues. In our experimental paradigm, depth cues other than shading are suppressed as far as possible. However, in these experiments there is one possibly disambiguating circumstance (or conflicting depth cue): the fact that stimuli are viewed on a computer screen, which presents—or at least is known to consist of—a well-nigh flat surface in the frontoparallel plane. The monitor screen may have constrained the mean or "global" slant of the settings, thereby fixing one of the global gauges. In fact, all observers reported that the images with real mean surface slant up to 20 deg looked like surfaces that were roughly in the frontoparallel plane. In natural viewing conditions the assumption of a particular mean slant may also be of importance.

  For all settings we compute the average angle of the normal settings with the observer-axis, a procedure that involves vector addition of the normal vectors. Results are displayed in Fig. 9. Clearly, surfaces are always set in the frontoparallel plane, except when strong ("almost") contour information is present in the image, which is the case only for the surfaces with slant 40 or 60 deg. In the latter case, only observer RE sets the average slant as large as the real global surface slant; the settings of the other observers always fall below the real slant for slanted surfaces. This is experimental evidence for an *interaction* of the two components of pictorial perception (Gregory, 1970); in this case, the frontoparallel screen surface and the depth relief from shading. A similar "regression to the picture plane" or "slant underestimation" has also been found for other depth cues such as texture and motion (see, e.g. Gibson, 1950; Braunstein, 1968).

  When a solution with global slant zero is chosen, there remains only one global gauge freedom. It is possible that in our experiments this freedom was fixed by an assumed light source direction (e.g. by constraining only the azimuth of the light source). It may also be that the visual system chooses the solution that is most cylindrical, since we have found more cylindrical points in the settings than in the real surfaces.
- In view of the great complexity and ambiguity of shape from shading, it is possible that observers invoke some simple trick to relate local luminance structure to local shape. Generally, the fact that only tiny fractions (around 0.1) of the variability of the settings can be accounted for by the local luminance structure pleads strongly against such tricks. Only very close to the occluding contours that are present in some images did we find that the tilt of the settings is in the direction of the luminance gradient. The main contribution to the depth structure that an occluding contour provides in our experiments is that it resolves the ambiguous *depth ordering* across the contour in a well-determined way. This phenomenon is well known in art history as the principle of vertical placement or height-in-the-field (Bunim, 1940): the higher side on the screen is seen as being behind. However, settings in the direction of the luminance gradient are certain to yield the real local surface orientation only *on* the occluding contour.
- Remaining options include Bayesian analysis regarding *a priori likely* surfaces underlying a given shading pattern, or demanding stability of the *a posteriori* surfaces under perturbation of, for example, the light source direction (Freeman, 1994). This type of analysis is attractive in view of the robustness of the results, but may not always be able to reduce the two-dimensional space of solutions for complex shading patterns to one *unique* surface. By working with random landscapes we have at least avoided suggesting likely surfaces. An investigation of the stability of the various possible solutions under changed lighting is beyond the scope of this paper. However, in the case of our images, luminance fields are non-degenerate, in the sense that spatial derivatives are generally non-zero. This means that also the shapes that correspond to these images are locally non-degenerate, or *generic*. Therefore, the solutions are expected to be equally stable, because catastrophes only occur in the case of degenerate solutions such as when solutions are flat in extended areas.

CONCLUSION

It is well known that predispositions and external information about the perceived object and the viewing conditions considerably influence the perception of surface relief from shading (see, e.g. Rittenhouse, 1786; Rock *et al.*, 1978). In this investigation we measure surface perception in the case of monocular shading information only, to the exclusion of as many external cues to the surface shape and viewing conditions as possible.

From the surface-normal adjustments on many different shaded landscapes, we have demonstrated that the probes are consistently set on "perceived" *surfaces*, and that this depth percept is mostly *stable*, despite a large theoretical ambiguity. Even though the mathematical solutions of the pure shape-from-shading problem form a continuous, two-parameter family of surfaces, observers choose discrete solutions, which are often found to be anti-correlated with and thus an *inversion* of the real

surface. For one observer, RE, inversions are even present within the set of measurements.

The resolution of the ambiguity must be due to internal *biases*. The net effect of these is to produce reasonably consistent, inter-subjectively correlating and in many cases veridical percepts. The biases were:

(1) A strong tendency to underestimate the global surface slant.
(2) In the case of a global slant in the settings, this slant is always in an upward direction. This corroborates results by Reichel and Todd (1990).
(3) If occluding contours are present in the shading pattern, the settings on the perceived foreground near the contours correspond to occluding "rims" on the surface which are oriented parallel to the picture plane.
(4) A less pronounced but clearly present bias for illumination from the right or overhead illumination.
(5) A scaling of the depth dimension—or, possibly, of the depth range as measured from an inferred light source direction. This range is always left unspecified in the shading pattern as it covaries with the relative strength of the light sources. The depth-range bias differs between observers.
(6) From the first-order variation in locally adjusted surface normals, one can deduce a slight bias towards locally flat—cylindrical—shapes.

It would be interesting to see how the visual system overcomes these biases when disambiguating information with respect to the depth structure is added to pure shading. Would the visual system show hysteresis and discrete, sudden transitions to other interpretations of the available cues to depth, such as are found for ambiguous two-dimensional figures (Attneave, 1971) and such as are seen in the few ambiguous series of settings that we encounter (those for observer RE, Fig. 4), or would it be possible to have continuous series of interpretations when, e.g. gradually more stereo information is added to the shading pattern? Brewster's informal results (1826), obtained by moving a second light source: "it is curious to observe the progress of the deception by which the depression is again changed into an elevation", indicate that the latter possibility should not immediately be ruled out (cf. also Bülthoff & Mallot, 1988).

Our differential-geometric approach to local shading, as adopted in the subsection Ecological Optics of Shape From Shading, may be useful in many other settings where shading is completely determined by the angle between the light vector and the surface normal ("generalized Lambertian shading"). This result generalizes one result by Oliensis (1991), who shows that if the light vector is known (thus if both degrees of freedom are fixed) pure shape from shading has a unique solution away from the contour. In machine-vision applications the light vector is often known (or else one is able to fix the gauge freedom in one of the other ways discussed above), such that our proposal of fitting second-order surfaces locally should yield a unique solution, which is at worst finitely degenerated.

REFERENCES

Attneave, F. (1971). Multistability in perception. *Scientific American, 225*, 62–71.

Braunstein, M. L. (1968). Motion and texture as sources of slant information. *Journal of Experimental Psychology, 78*, 247–253.

Brewster, D. (1826). On the optical illusion of the conversion of cameos into intaglios and of intaglios into cameos, with an account of other analogous phenomena. *Edinburgh Journal of Science, 4*, 99–108.

Bülthoff, H. H. & Mallot, H. A. (1988). Integration of depth modules: stereo and shading. *Journal of the Optical Society of America A, 5*, 1749–1758.

Bülthoff, H. H. & Mallot, H. A. (1990). Integration of stereo, shading and texture. In Blake, A. & Troscianko, T. (Eds), *AI and the eye*. New York: Wiley.

Bunim, M. S. (1940). *Space in medieval painting and the forerunners of perspective*. New York: Columbia University Press.

Burke, W. L. (1985). *Applied differential geometry*. Cambridge: Cambridge University Press.

Erens, R. G. F. & de Haan, E. (1993). Detection of second-order luminance structure. *Utrecht Biophysics Research Institute Technical Report*, UBI-T.92.MF-045.

Fisher, N. I. & Lee, A. J. (1983). A correlation coefficient for circular data. *Biometrika, 70*, 327–332.

Freeman, W. T. (1994). The generic viewpoint assumption in a framework for visual perception. *Nature, 368*, 542–545.

Gibson, J. J. (1950). The perception of visual surfaces. *American Journal of Psychology, 63*, 367–384.

Gregory, R. L. (1970). *The intelligent eye*. London: Weidenfeld and Nicholson.

von Helmholtz, H. (1867). *Handbuch der Physiologischen Optik*. Leipzig: Voss.

Hettmansperger, T. P. (1984). *Statistical inference based on ranks*. New York: Wiley.

Horn, B. K. P. (1990). Height and gradient from shading. *International Journal of Computer Vision, 5*, 37–75.

Koenderink, J. J. (1984). What does the occluding contour tell us about solid shape? *Perception, 13*, 321–330.

Koenderink, J. J. (1990). *Solid shape*. Cambridge, Mass.: MIT Press.

Koenderink, J. J. & van Doorn, A. J. (1987). Representation of local geometry in the visual system. *Biological Cybernetics, 55*, 367–375.

Koenderink, J. J. & van Doorn, A. J. (1992). Receptive field assembly pattern specificity. *Journal of Visual Communication and Image Representation 3*, 1–12.

Koenderink, J. J. & van Doorn, A. J. (1993). Illuminance critical points on generic smooth surfaces. *Journal of the Optical Society of America A, 10*, 844–854.

Koenderink, J. J., van Doorn, A. J. & Kappers, A. M. L. (1992). Surface perception in pictures. *Perception & Psychophysics, 52*, 487–496.

Mingolla, E. & Todd, J. T. (1986). Perception of solid shape from shading. *Biological Cybernetics, 53*, 137–151.

Oliensis, J. (1991). Shape from shading as a partially well-constrained problem. *CVGIP: Image Understanding, 54*, 163–183.

Pentland, A. P. (1982). Finding the illuminant direction. *Journal of the Optical Society of America, 72*, 448–455.

Pfuhl, E. (1923). *Malerei und Zeichnung der Griechen*. Munich: Bruckmann. *Vide* Index III B 7, "Licht und Schatten".

Regan, D. & Hamstra, S. J. (1992). Shape discrimination and the judgement of perfect symmetry: Dissociation of shape from size. *Vision Research, 32*, 1845–1864.

Reichel, F. D. & Todd, J. T. (1990). Perceived depth inversion of smoothly curved surfaces due to image orientation. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 653–664.

Rittenhouse, D. (1786). Explanation of an optical deception. *Transactions of the American Philosophical Society, 2,* 37–42.

Rock, I., Shallo, J. & Schwartz, F. (1978). Pictorial depth and related constancy effects as a function of recognition. *Perception, 7,* 3–19.

Stevens, K. A. (1981). The visual interpretation of surface contours. *Artificial Intelligence, 17,* 47–73.

Stevens, K. A. (1983). Slant–tilt: The visual encoding of surface orientation. *Biological Cybernetics, 46,* 183–195.

Todd, J. T. (1985). Perception of structure from motion: Is projective correspondence of moving elements a necessary condition? *Journal of Experimental Psychology: Human Perception and Performance, 11,* 689–710.

Todd, J. T. & Reichel, F. D. (1989). Ordinal structure in the visual perception and cognition of smoothly curved surfaces. *Psychological Review, 96,* 643–657.

Verbraeken, R. (1979). *Clair-Obscur–histoire d'un mot.* Libraire des Arts et Métiers, Nogent-le-Roi, France.

Acknowledgements—We would like to thank Susan te Pas for acting as observer in the experiments. E.H. was funded by the Biophysics Foundation of the Netherlands Organization for Scientific Research (N.W.O.), R.E. by the SPIN project "3D Computer Vision" of the Dutch Ministry of Economic Affairs and A.N. by SNN, the Dutch Foundation for Neural Network Research.